

⑫ 特 許 公 報 (B2)

平4-2198

⑤Int. Cl.³
G 10 L 3/00識別記号 庁内整理番号
3 0 1 D 8842-5H

⑭公告 平成4年(1992)1月16日

発明の数 1 (全7頁)

⑮発明の名称 音声認識方法

⑯特 願 昭58-165547

⑰公 開 昭60-57395

⑱出 願 昭58(1983)9月8日

⑲昭60(1985)4月3日

⑳発 明 者 井 上 郁 夫 神奈川県川崎市多摩区東生田3丁目10番1号 松下技研株式会社内

㉑発 明 者 二 矢 田 勝 行 神奈川県川崎市多摩区東生田3丁目10番1号 松下技研株式会社内

㉒発 明 者 藤 井 諭 神奈川県川崎市多摩区東生田3丁目10番1号 松下技研株式会社内

㉓発 明 者 森 井 秀 司 神奈川県川崎市多摩区東生田3丁目10番1号 松下技研株式会社内

㉔出 願 人 松下電器産業株式会社 大阪府門真市大字門真1006番地

㉕代 理 人 弁理士 小 鍛 治 明 外2名

審 査 官 中 村 和 男

㉖参 考 文 献 特開 昭56-29293 (JP, A) 特開 昭56-140398 (JP, A)
特開 昭59-176793 (JP, A)

1

2

⑳特許請求の範囲

1 概念的につながりを持つ n 種の単語辞書を備え、一時に続けて発声された n 単語の入力音声を用いて単一の単語を認識するに際し、第 i 番目の入力単語($i=1, 2, \dots, n-1$)に対して第 i 番目の辞書を用いて事後確率化した距離尺度を用いて第 m 候補までの類似度を求め、第 $i+1$ 番目の入力単語に対しては第 i 番目における第1～第 m 候補に概念的に接続する辞書中の項目を第 $i+1$ 番目の辞書から選択して、これらの辞書中の項目に対して第 m' 候補までの類似度を求め、このような手続きによつて n 個の入力単語全てに対する候補単語とその類似度を求めた後、概念的に接続第1～第 n 入力音声に対する候補単語の類似度の合計を計算して、類似度の合計を最大とする単語辞書中の項目を認識結果とすることを特徴とする音声認識方法。

発明の詳細な説明

産業上の利用分野

本発明は人間の発声した音声に応じた動作を機械に行なわせることを可能とする音声認識方法に関するものである。

従来例の構成とその問題点

人間の声を認識する音声認識装置は、計算機やワードプロセッサの入力、音声による品物の仕分け作業等のいろいろな分野に於て、円滑に作業を行なう為の有効な入力手段として注目を集めている。

10 従来の単語音声認識装置では、単独発声された単語音声をマイクロホンより入力し、音響分析を行なつた後特徴パラメータを抽出し、単語辞書として格納されている単語標準パターンとの比較を行ない、確からしさの度合(尤度)の最も高い単語を認識結果として出力していた。しかしこの様な方法では認識対象語いを多くすると、辞書中に類似した単語が含まれる割合が増えて来る結果、認識率の急激な低下をまねく為、実用的なレベルの認識率を確保するには認識対象語いを小数に限

定するか、装置自体に応答機能を持たせて誤りである場合には次に尤度の高い単語を出力するか、あるいは複数の候補を出力して選択させるといった方法をとらざるを得ず、用途が限定されたり、手間がかかるといった問題があった。

発明の目的

本発明は、従来の単語音声入力装置における前記の問題点を解決し、多数語いが扱え、高い認識率を有しなおかつオペレーションが簡単であるという特徴を有する、より実用的な音声認識装置の実現を可能とする音声認識方法の提供を目的とする。

発明の構成

本発明は上記目的を達成するためになされたもので、概念的につながりを持つ n 種の単語辞書を備え、一時に続けて発声された n 単語の入力音声を用いて単一の単語（多くの場合第 n 番目の単語辞書中の項目）を認識するに際し、第 i 番目の入力単語（ $i = 1, 2, \dots, n-1$ ）に対して第 i 番目の辞書を用いて事後確率化した距離尺度を用いて第 m 候補までの類似度を求め、第 $i+1$ 番目の入力単語に対しては第 i 番目における第 $1 \sim$ 第 m 候補に概念的に接続する辞書中の項目を第 $i+1$ 番目の辞書から選択して、これらの辞書中の項目に対して第 m' 候補までの類似度を求め、このような手続きによつて n 個の入力単語全てに対する候補単語とその類似度を求めた後、概念的に接続する第 $1 \sim$ 第 n 入力音声に対する候補単語の類似度の合計を計算して、類似度の合計を最大（または累積距離値を最小）とする単語辞書中の項目を認識結果とすることを特徴とする音声認識方法を提供するものである。

実施例の説明

我々が、多くの物の中からある特定の物を探す場合、ただ片端から探すのでは非常に時間がかかってしまう。この様な場合、個々の物がそれぞれが持っている属性を基に系統的に分類整理されていれば、ある特定の物のもつ属性をより大きな分類から順に辿っていくことにより、容易に捜し出すことが出来る。

単語音声認識装置に於ても、この様な方法で認識を行えばより効率的な認識が行なえるだけでなく、より確実な認識結果を得ることが可能である。

大語いを扱う単語音声認識装置の場合、多くの単語の中からある 1 つの単語をより正確に認識することを最大な目的とするならば、前述の様な方法を用いることによりこれを実現することが可能である。例えば、日本の都市名单語を認識対象とした場合、その都市の属する区や市毎に 1 つの都市名单語集団とし、それぞれの都市の属する市や区の名前を集めて市（区）名单語集団とし、同様にしてそれぞれの区や市の属する都道府県名を集めて都道府県名单語集団とする tree 型の階層構造を作成しておき、一回の入力で例えば県名、市（区）名、都市名の順に 3 語を区切つて発声し、認識装置ではこの順序でそれぞれの階層に於ける認識候補を複数求めておき、それぞれの階層の認識候補相互の tree 上の位置関係と、それぞれの単語に対する尤度を基に都市名の決定を行なうことにより、認識率の大幅な低下を伴うことなくして大語いを対象とした単語の認識が可能となる。また、この様な方法を用いれば、発声は一度の入力で行なえるので、質問応答を主体とした従来の方法の様なめんどろな手続きも不要である。

ここで、単語の認識に用いる尤度の値は以下の様にして求めることが出来る。

Ω を単語辞書、 D を辞書項目とすると、入力情報 x が与えられた時の認識結果 D^* は、入力情報 x が与えられた時の辞書項目 D の事後生起確率が最大となる D によつて与えられる。この関係は、Bayes の定理により

$$P(D^* | x) = \max_{D \in \Omega} P(D | x) \\ = \max_{D \in \Omega} \frac{P(D)P(x | D)}{P(x)}$$

と表わされる。 $P(D)$ は一定としてよく、 $P(x)$ は D に関わらず共通であるから、結局

$$S(D^* | x) = \max_{D \in \Omega} P(x | D)$$

を満足する D を求めればよく、その時の $P(x | D)$ の値が辞書項目に対する尤度 S となる。

前記説明で述べた様に、本発明は 1 回の入力で複数の単語を発声しておき、その第 1 語目で第 1 の属性（例えば県名）の中から認識し、第 2 語目で第 1 語目の認識結果に代表される第 2 の属性（例えば市名）の中から認識し、第 3 語目で第 2 語目の認識結果に代表される第 3 の属性（例えば

区名)の中から認識するという様に、階層的に認識を行なつてゆくものである。

以下に本実施例の詳細を図面を用いて説明する。

第1図は本実施例の認識に用いる単語辞書treeの構成の例を示す図である。第2図は、本実施例による単語音声認識装置の構成の例を示す図である。第1図及び第2図を参照しながら、都市名单語を対象として認識を行なう場合を例にとり本実施例を説明する。

第1図に於いて、1~5はそれぞれ単語辞書を表わし、6は第1層の単語辞書列、7は第2階層の単語辞書列、8は第3階層の単語辞書列を表わす。各単語辞書は、より上位の階層に於ける単語が示す属性を持つ単語名から成る。認識されるべき都市名单語は第3階層の単語辞書列8にあり、第2階層の単語辞書列7にはそれぞれの都市が属する市(区)名から成る単語辞書が、さらに、第1階層の単語辞書列6にはそれぞれの市(区)が属する都道府県名から成る単語辞書が入っている。

都市名单語の認識を行なう際、1回の入力で、その都市が属する県名、市(区)名、その都市名の順に3語に区切りながら発声し、それぞれの階層毎に入力単語と辞書単語との間でマッチングを行ない、その結果得られた事後確率化された尤度(明細書第6頁第7行目から第7頁第2行目に記載の高いものから順に認識候補とする。

先ず説明をわかりやすくするために、最も単純な場合としてtreeの上の層から第1認識の枝を順に辿り、認識都市名を出力するものとし、都市名/KANDA(神田)/を認識する為に、/TOOKYOOTO(東京都)/CHIYODAKU(千代田区)/KANDA(神田)/と1語ずつ区切って発声する場合について説明する。

第2図に於て、初期状態では階層計数部13の値は1となっており、単語辞書選択部12では第1階層の単語辞書列6の都道府県名辞書が選択される。音声入力部9から第1語目が入力されると、単語認識部10では、単語辞書選択部12で選択された都道府県名辞書が単語辞書群格納部11から取り出され、その中の1語(例えば東京都)が認識され、その結果が単語判定部14と単語辞書選択部12へ送られる。この際、階層計数

部の値は+1されて、次の認識が第2階層の単語辞書列7に対するものであることを示す。単語選択部12では、第1語目の認識結果と階層計数部13の値によつて次に入力されるべき単語の属する単語辞書が選択される。音声入力部9から第2語目が入力されると、単語認識部10では、単語辞書選択部12で選択された第2階層の単語辞書列7の区名辞書が単語辞書群格納部11から取り出され、その中の1語(例えば千代田区)が認識され、その結果が単語判定部14と単語辞書選択部12に送られる。この際、階層計数部13の値は更に+1されて、次の認識が第3階層の単語辞書列8に対するものであることを示す。単語辞書選択部12では、第2語目の認識結果と階層計数部10の値によつて次に入力されるべき単語の属する単語辞書が選択される。音声入力部9に第3語目が入力されると、単語認識部10では、単語辞書選択部12で選択された第3階層の単語辞書列8の都市名辞書が単語辞書群格納部11から取り出され、その中の1語(例えば神田)が認識され、その結果が単語判定部14へ送られる。この様にして全ての階層についての単語認識が終わると、単語判定部14には/東京都/千代田区/神田/という認識結果が得られ、これらの全てあるいは一部が最終的な認識結果として出力される。この場合、/KANDA/が認識される為には、第1階層の単語辞書列で/TOOKYOOTO/、第2階層の単語辞書列で/CHIYODAKU/、第3階層の単語辞書列で/KANDA/がそれぞれ正しく認識されなければならない。しかし、各階層に於ける単語辞書の辞書項目数は、全ての都市名を1つの単語辞書とする従来の辞書項目数に比べて遥かに少ないので、各階層に於ける単語認識率は非常に高くなる為最終的な単語認識率であるそれらの積の値も高くなる。なお、本実施例のように各単語辞書の辞書項目数を限定することにより誤認識の起きる割合を十分少なくする場合には、単語判定部14は必ずしも必要でない。入力音声は3単語で、単語辞書列が3層の場合で認識候補数を1つに絞つた時の単語認識フローチャートを第3図に示す。しかし、このような第1位の認識結果を用いる方法では、総合認識率はそれぞれの認識率の積になるので、あまりよい結果は期待できない。次に本発明の主眼である、各階層

に於て、認識単語候補の数を1つに限定せずに数番目の候補まで求める方法を説明する。複数候補を用いることによりそれらの中に正解単語が含まれる確率は更に高まり、総合認識率は高くなる。この場合、第2図の例で第1語目の都道府県名单語を認識されるとき、単語認識部10では、入力単語と都道府県名单語辞書の辞書項目との事後確率化された尤度の高いものから例えば第3位までが認識候補として認識され、その結果が単語判定部14及び単語辞書選択部12に送られる。単語辞書選択部12では、第3位までの認識候補に対応する3つの市(区)各単語辞書が選択される。次に、第2語目の市(区)各単語が認識されるとき、単語認識部10では、3つのそれぞれの市(区)名单語辞書について、入力単語と辞書項目との事後確率化された尤度の高いものから例えば第3位までが認識候補として単語判定部14及び単語辞書選択部12に送られる。単語辞書選択部12では、それぞれの認識候補に対する計9つの都市名单語辞書が選択される。最後に、第3語目の都市名单語を認識されるとき、単語認識部10では、9つのそれぞれの都市名单語辞書について、入力単語と辞書項目との尤度の最も高いものが認識候補として単語判定部14に送られる。この様にして全ての階層についての単語認識が終わると、単語判定部14では、都道府県名、市(区)名、都市名と連なる9通りの単語tree系列の組み合わせができる。ここで、第1語目の入力に対する辞書項目の事後確率化された尤度を大きいものから順に S_i ($i = 1, 2, 3$)とし、第2語目の入力に対する i 番目の辞書の辞書項目の事後確率化された尤度を大きいものから順に S_{ij} ($j = 1, 2, 3$)とし、第3語目の入力に対する j 番目の辞書の辞書項目の事後確率化された尤度が最大のものを S_{ij} とすると、単語判定部14での判定規則は、例えば9通りの単語tree系列の各系列毎の事後確率化された尤度と L

$$L = S_i + S_{ij} + S_{ij}$$

$$(i = 1, 2, 3; j = 1, 2, 3)$$

あるいはそれぞれの辞書項目数によつて重みづけられた事後確率化された尤度と L_w

$$L_w = \omega_i S_i + \omega_{ij} S_{ij} + \omega_{ij} S_{ij}$$

$$(i = 1, 2, 3; j = 1, 2, 3)$$

$$(\omega_i, \omega_{ij}, \omega_{ij} \text{は各階層に於ける辞書項目数に}$$

よる重み)が最も大きくなる単語tree系列の全単語あるいはその中の都市名单語を認識結果とすればよい。この様に、各階層に於ける認識候補を複数にし、各階層での事後確率化された尤度値から、単語tree系列の単位で認識結果を決定することにより、上位の階層に於ける認識の段階で正解単語に至るtreeの枝の脱落が起こるのを防ぎ、更に認識率を向上させることができる。

入力音声が入力単語、単語辞書列が3層構成で認識候補数を複数許す場合の単語認識フローチャートを第4図に示す。この場合認識候補数は第1番目の入力単語には一律に m 個、第2番目の入力単語には一律に n 個と定めた場合について示した。本実施例において、例えば第1層から第3層までの単語辞書列の各単語辞書がそれぞれ100単語から成り、特別な認識傾向を持たない一般的なものである場合を仮定すれば、不特定話者を対象とした100単語の認識率が96.6%の単語音声認識装置では、100万単語について90%以上の認識率を得ることが可能となる。また、この時の階層構造化の為に新たに作成する単語辞書の量は1.1%の増加にすぎず、十分実用化可能である。また、辞書項目の標準パターンと入力パターンとのマッチング演算の回数は、各単語辞書の辞書項目数が全て n で3層から成る場合、階層構造にして第1候補のみを選ぶ場合と階層構造としない場合との比はほぼ $3n : n^3$ となり本実施例を用いることにより演算時間を大幅に削減することができる。

発明の効果

以上要するに本発明は、概念的につながりを持つ n 種の単語辞書を備え、一時に続けて発声された n 単語の入力音声を用いて単一の単語(多くの場合第 n 番目の単語辞書中の項目)を認識するに際し、第 i 番目の入力単語($i = 1, 2, \dots, n-1$)に対して第 i 番目の辞書を用いて事後確率化した距離尺度を用いて第 m 候補までの類似度を求め、第 $i+1$ 番目の入力単語に対しては第 i 番目における第1~第 m 候補に概念的に接続する辞書中の項目を第 $i+1$ 番目の辞書から選択して、これらの辞書中の項目に対して第 m' 候補までの類似度を求め、このような手続きによつて n 個の入力単語全てに対する候補単語とその類似度を求めた後、概念的に接続する第1~第 n 入力音声に対する候補単語の類似度の合計を計算して、類似

度の合計を最大（または累積距離値を最小）とする単語辞書中の項目を認識結果とすることを特徴とする音声認識方法を提供するもので、数十から数百単語を1つの単語辞書として階層化を行なう事により、新たに作成する辞書項目の数の増加を数%以下におさえ、2層の構成の場合は数千語から数万語の単語が、また3層構成の場合は数万語から数百万語の単語を高い認識率で認識することが可能となり、従来、実用化にならなかった大語い単語音声認識装置が簡単に実現でき、また大語いの場合、入力パターンと単語辞書の標準パターンとのマッチングを全辞書項目に対して行なわず

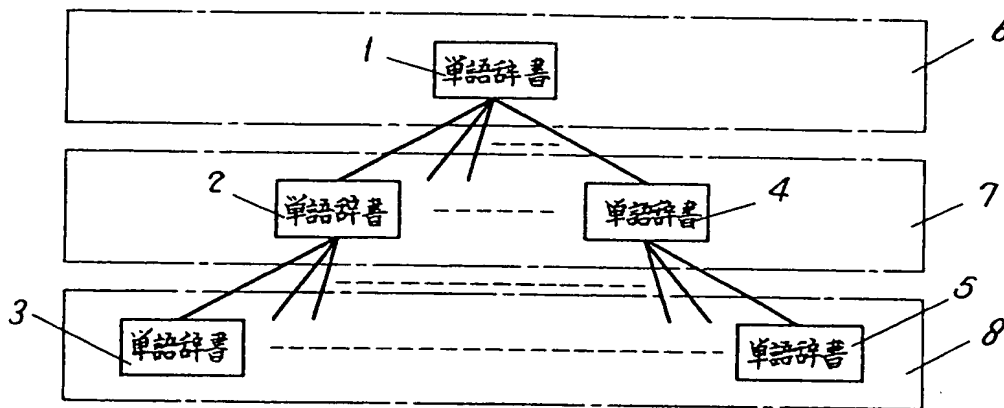
に済む為演算時間が大幅に短縮される。

図面の簡単な説明

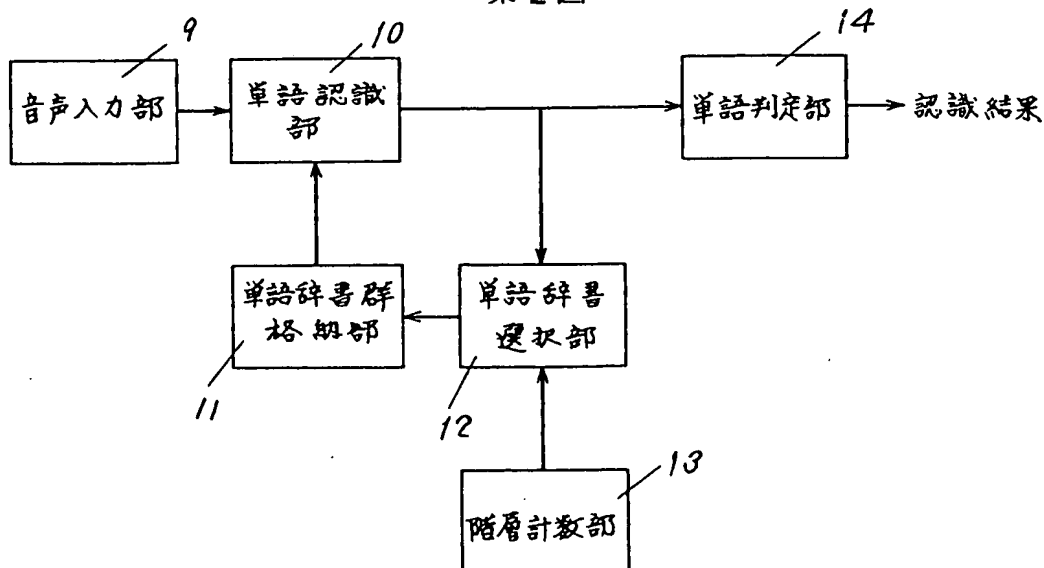
第1図は本発明の単語辞書treeの構成の例を示す図、第2図は本発明の一実施例である単語音声認識装置のブロック図、第3図及び第4図は本実施例のフローチャートを示す図である。

1～5……単語辞書、6……第1階層の単語辞書列、7……第2階層の単語辞書列、8……第3階層の単語辞書列、9……音声入力部、10……単語認識部、11……単語辞書群格納部、12……単語辞書選択部、13……階層計数部、14……単語判定部。

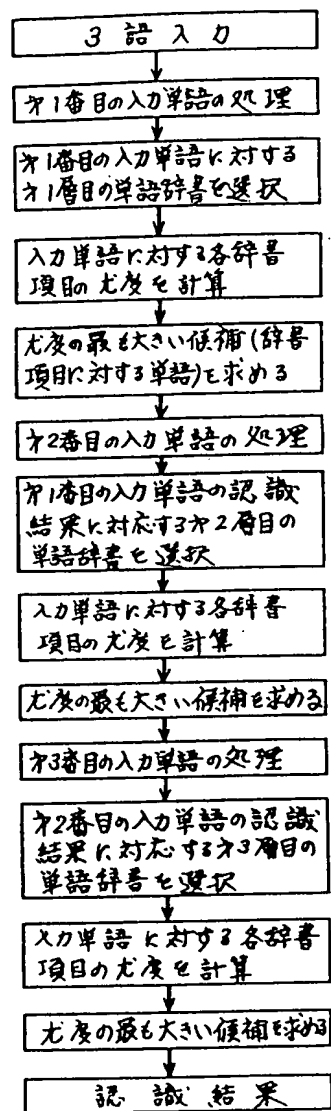
第1図



第2図



第 3 図



第 4 図

